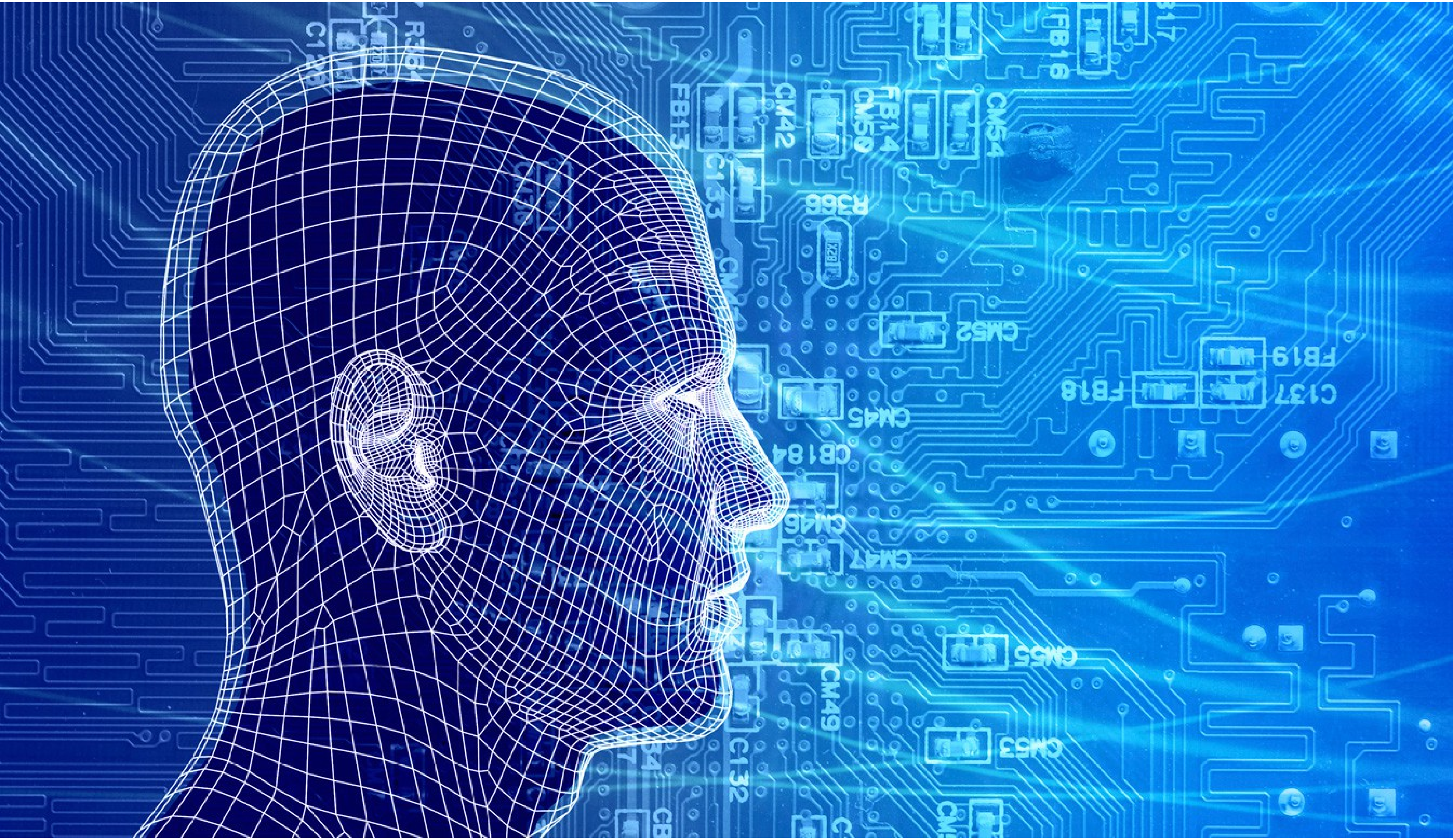
Can machines think like human beings and beyond!



Time Series Forecasting

ARIMA models

- **A**uto-**R**egressive **I**ntegrated **M**oving **A**verage
- Are an adaptation of discrete-time filtering methods developed in 1930's-1940's by electrical engineers (Norbert Wiener et al.)
- Statisticians George Box and Gwilym Jenkins developed systematic methods for applying them to business & economic data in the 1970's (hence the name "Box-Jenkins models")

What ARIMA stands for

- A series which needs to be differenced to be made stationary is an “integrated” (**I**) series
- Lags of the stationarized series are called “auto-regressive” (**AR**) terms
- Lags of the forecast errors are called “moving average” (**MA**) terms

ARIMA forecasting equation

- Let Y denote the *original* series
- Let y denote the *differenced* (stationarized) series

No difference (d=0): $y_t = Y_t$

First difference (d=1): $y_t = Y_t - Y_{t\text{-}1}$

Second difference (d=2): $y_t = (Y_t - Y_{t\text{-}1}) - (Y_{t\text{-}1} - Y_{t\text{-}2})$

$$= Y_t - 2Y_{t\text{-}1} + Y_{t\text{-}2}$$

Note that the second difference is not just the change relative to two periods ago, i.e., it is *not* $Y_t - Y_{t\text{-}2}$. Rather, it is the change-in-the-change, which is a measure of local "acceleration" rather than trend.

P,d,q parameters

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA(p,d,q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

- **p**: The number of lag observations included in the model, also called the lag order.
- **d**: The number of times that the raw observations are differenced, also called the degree of differencing.
- **q**: The size of the moving average window, also called the order of moving average.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.

Forecasting equation for *y*

$$\hat{y}_t = \underbrace{\mu}_{\text{constant}} + \underbrace{\phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}}_{\text{AR terms (lagged values of } y\text{)}}$$

$$\underbrace{- \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}}_{\text{MA terms (lagged errors)}}$$

By convention, the AR terms are + and the MA terms are –

Understanding prediction in ARIMA

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

where, $Y *{t-1}$ is the lag1 of the series, $\beta* 1$ is the coefficient of lag1 that the model estimates and α is the intercept term, also estimated by the model.

Likewise a pure **Moving Average (MA only) model** is one where Yt depends only on the lagged forecast errors.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations :

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + .. + \beta_0 Y_0 + \epsilon_{t-1}$$

ARIMA prediction Summarized

ARIMA model in words:

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

VECTOR AUTOREGRESSION (VAR)

- Vector autoregressive models (VARs) are used to deal with forecasting two or more time series.
- In VAR we have one equation for each variable and each equation contains only the lagged values of that variable and the lagged values of all other variables in the system.
  - As in the case of the univariate time series, in VAR we also require the time series to be stationary.
  - If each variable in the VAR is already stationary, each equation in it can be estimated by OLS.
  - If each variable is not stationary, we can estimate VAR only in the first-differences of the series.
  - If individual variables in VAR are nonstationary, but are cointegrated, we can estimate VAR by taking into account the error correction term, which is obtained from the cointegrating regression.
  - This leads to **vector error correction model (VECM)**.

NATURE OF CAUSALITY

- VAR modes can also be used to shed light on causality among variables.
- The basic idea behind VAR causality testing is that the past can cause the present and the future, but not the other way around.
- Granger causality: In establishing causality, we must make sure that the underlying variables are stationary. If they are not, we have to difference the variables and run the causality test on the differenced variables.
- However, if the variables are nonstationary, but are integrated, we need to use the error correction term to account for causality, if any.

GRANGER CAUSALITY TEST

- 1. Regress current *Y* on all lagged *Y* terms and other variables, if any (such as trend), but do not included the lagged *X* terms in this regression. This is the restricted regression. Obtain the restricted residual sum of squares, *RSSr*.
- 2. Reestimate the equation including the lagged *X* terms. This is the unrestricted regression. From this regression obtain the unrestricted residual sum of squares, *RSSur*.
- 3. The null hypothesis is that the lagged *X* terms do not belong in the regression.
- 4. To test the null hypothesis, we apply the *F* test, which is:

$$F = \frac{(RSS_r - RSS_{ur})/m}{RSS_{ur}/(n-k)}$$

which has m and $(n-k)$ df, where m is the number of lagged *X* terms, k is the number of parameters estimated in the unrestricted regression, and n is the sample size.

- 5. If the computed *F* value exceeds the critical *F* value at the chosen level of significance, we reject the null hypothesis.

GRANGER CAUSALITY TEST (CONT.)

- Run the test again switching X and Y. There are four cases:
- 1. *Unidirectional causality from X to Y*
- 2. *Unidirectional causality from Y to X*
- 3. *Feedback or Bilateral causality* is indicated when the sets of Y and X coefficients are statistically significantly different from zero in both regressions.
- 4. *Independence* is suggested when the sets of Y and X coefficients are not statistically significant in either of the regressions.

Hypothesis Test

Hypothesis Testing

Hypothesis testing is an inferential statistical technique that determines if a certain condition is true for the population.

Alternative Hypothesis (H1)	Null Hypothesis (H0)
A statement that has to be concluded as true.	A statement of "no effect" or "no difference".
It's a research hypothesis.	It's the logical opposite of the alternative hypothesis.
It needs significant evidence to support the initial hypothesis.	It indicates that the alternative hypothesis is incorrect.
If the alternative hypothesis garners strong evidence, reject the null hypothesis.	Weak evidence of alternative hypothesis indicates that the null hypothesis has to be accepted.

Hypothesis Testing – Error Types

Representation of decision parameters using null hypothesis

Type I Error (α)	• Rejects the null hypothesis when it is true • The probability of making Type I error is represented by α
Type II Error (β)	• Fails to Reject the null hypothesis when it false • The probability of making Type II error is represented by β
p-value	• The probability of observing extreme values • Calculated from collected data

Decision	Ho is True	Ho is False
Fail to Reject Null	Correct	Type II Error
Reject Null	Type I Error	Correct

Hypothesis Testing - Process

There are four steps to the hypothesis testing process.

Step 1

Set Hypothesis
H0 ($\mu_1 = \mu_2$) : Equality
H1($\mu_1 \neq \mu_2$) : Difference

Step 2

Set α Significant Level
Set α or choose the significant level for the population.

Step 3

Collect Data
Collect sample from population

Step 4

Make decision
p-value < α
p-value ≥ α

Reject the null hypothesis if p-value < α
Fail to reject the null hypothesis if p-value ≥ α

Chi-Square Test

It is a hypothesis test that compares the observed distribution of your data to an expected distribution of data.

Test of Association:
To determine whether one variable is associated with a different variable. For example, determine whether the sales for different cellphones depends on the city or country where they are sold.

Test of Independence:
To determine whether the observed value of one variable depends on the observed value of a different variable. For example, determine whether the color of the car that a person chooses is independent of the person's gender.

Test is usually applied when there are two categorical variables from a single population.

Chi Square Test - Example

An example of Chi-Square test.

Null Hypothesis

- There is no association between gender and purchase.
- The probability of purchase does not change for 500 dollars or more whether female or male.

Alternative Hypothesis

- There is association between gender and purchase.
- The probability of purchase over 500 dollars is different for female and male.

	<$500	>$500
fo	.55	.45
fo	.75	.25

Types of Frequencies

Expected and observed frequencies are the two types of frequencies.

Expected Frequencies (fe)

The cell frequencies that are expected in a bivariate table if the two tables are statistically independent.

Observed Frequencies (fo)

- There is association between gender and purchase.
- The probability of purchase over 500 dollars is different for female and male.

	Purchases	
	<$500	>$500
fo	.55	.45
fo	.75	.25

No Association
Observed Frequency = Expected Frequency

Association
Observed Frequency ≠ Expected Frequency

Features of Frequencies

The formula for calculating expected and observed frequencies using Chi Square:

$$\sum \frac{(f_e - f_o)^2}{f_e}$$

Features of Expected and Observed frequencies:

- Requires no assumption of the underlying population
- Requires random sampling

Thank You!